Dementia Patients Analysis

Chenyang Pan
1005131554

1. Introduction

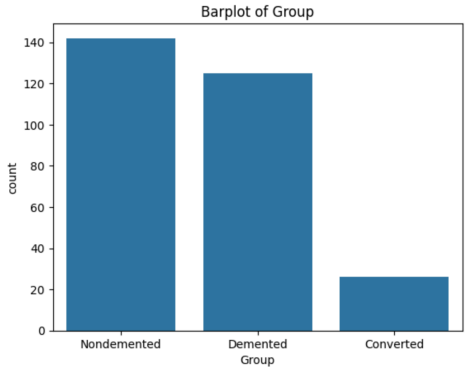
Dementia is a widespread disease mainly occurring in the orders group. Until today, there is none of the treatments which can effectively cure dementia. In this analysis,our goal is to investigate how cognitive function may be affected by dementia over time through mini state mental examination tests and how normalized total brain volume is affected. In order to achieve this goal, we compare the changes in MMSE and nWBV score from first visit to second visit based on three different groups, people with dementia, people without dementia, people who initially do not have dementia but had it in later visits. A in depth analysis is applied to a data set found on Kaggle which is about longitudinal study on MRI results of patients without or without dementia.

2. Data cleaning

16 features are initially found when we have a first look at the first several rows of the dataset. Since our goal in this analysis is to explore how the MMSE score and nWBV change over time, from first visit to second visit through comparing demented group, nondemented group and converted group, these only 5 features are left to make further analysis, which are the subject ID, group, Visit, MMSE and nWBV. Moreover, in the data cleaning section, the values in Visit are renamed to First visit or Second visit depending on whether the original value is 1 or 2. Doing so, we get better understanding and clarity on the meaning of the values in the Visit variable.

3. Exploratory Data Analysis

As the first step of Exploratory Data Analysis (EDA), we first count numerically the number of first visits and number of the second visits and find that there are 150 patients who have gone to the first visit and 143 patients who have gone to the second visit. This data tells us that some of the patients only visit once. Then a bar plot distribution of the dementia group is shown on the right. We can see that the number of nondemented patients has a similar number than demented patients but being slightly higher.



Next two summary statistics are shown in the tables below.

The table below shows the summary statistic for MMSE score for patients in different groups, demented, nondemented and converted. There are some interesting findings from this summary. We can detect that the nondemented group has the highest mean MMSE score followed by the converted group. Also the nondemented group has the lowest standard deviation, IQR and spreading indicating the MMSE score in thai group is more stable than the other two.

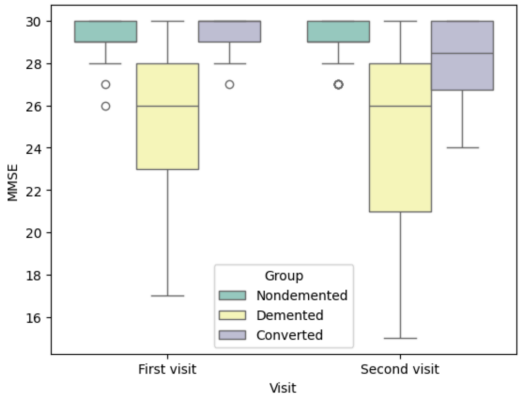
	Min	Mean	Max	Spread	Median	Standard deviation	IQR
Demented	15	24	30	15	26	3.908	6
Nondemented	26.0	29.155	30.0	4.0	29.0	0.902	1.0

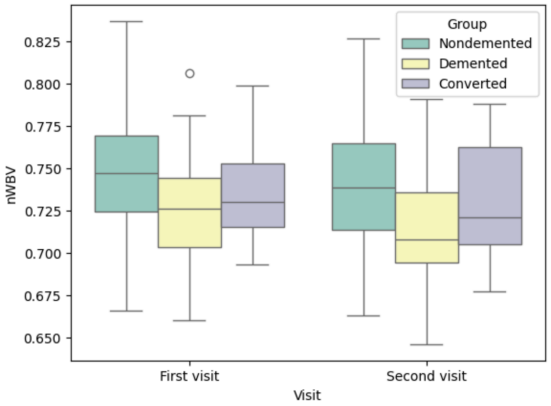
Converted	24.0	28.731	30.0	6.0	29.0	1.687	2.0

The below table shows the summary statistics for nWBV score for patients in different groups, demented, nondemented and converted. Similarly finding can be found. The nondemented group has the highest mean for nWBV. However, three groups have similar standard deviation, spread and IQR.

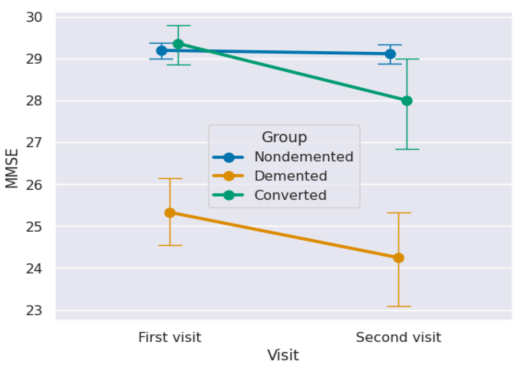
	Min	Mean	Max	Spread	Median	Standard deviation	IQR
Demented	0.646	0.719	0.806	0.160	0.715	0.033	0.046
Nondemented	0.663	0.742	0.837	0.17	0.739	0.039	0.053
Converted	0.677	0.733	0.799	0.122	0.725	0.034	0.048

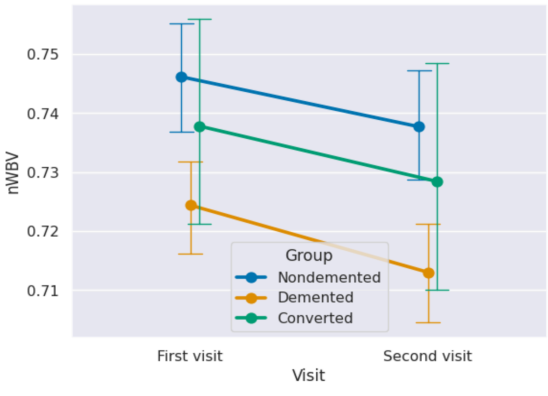
Next, two side by side boxplots are made so that we can visually compare the MMSE score and nWBV at first visit and second visit based on each dementia group. The bottom left boxplots indicating the MMSE score of different groups at the two visits, we can find an obvious distinction on the demented group in both visits compared to the other two groups, that is, a significant lower median and larger spread is observed. On the bottom left boxplots indicating the nWBV at first second visit based on each group. Unlike the MMSE boxplots, we do not see a large distinction across all groups but we can still see that the demented group has the lowest median compared to the other two.





Lastly I made two pointplot showing the changes over time of MMSE and nWBV from the first visit to the second visit separated the three groups. The bottom left point plot shows the changes of MMSE for different groups over time. We can see that there is a significant difference in rate of change between demented/converted group and nondemented group. A steep negative slope can be observed for demented/converted group and an almost flat slope is found on the nondemented group. On the bottom right shows the nWBV plot, unlike MMSE, the slope across three groups shows to be in a similar rate of decay, but the non demented group shows to be slightly slower decay in nWBV.





4. Two Way Mixed Effect ANOVAs

4.1 Model 1: MMSE

Null hypothesis: 1, there is no significant difference in MMSE score across two visits for all patients. 2, there is no significant difference in MMSE score across all dementia groups.3, The changes in MMSE score over two visits is the same for all groups.
Alternative hypothesis: 1, there is significant difference in MMSE score across two visits for all patients. 2, there is a significant difference in MMSE score across all dementia groups. 3, The changes in MMSE score over two visits differ across groups.

4.1.1 Model Result

The model result shows below in the ANOVA table, we can see that the sum of squared for group effect is 1328.421 which is the largest. A p value of smaller than 0.00, which is smaller the level of significance indicating that we have strong evidence to reject the null hypothesis that there is no significant difference in MMSE score across all dementia groups. Similar in for Visit effect and interaction effect, the p values are 0.003 and 0.037 which are both smaller than 0.05, thus , we reject all null hypotheses.

	SS	DF1	DF2	MS	F	P value
Group	1328.421	2	140	664.211	56.212	<0.001
Visit	22.378	1	140	22.378	8.859	0.003
Interaction	17.000	2	140	8.500	3.365	0.037

In the result of the post hoc test, we found that whether in the first visit or second visit, we have detected a significant difference in MMSE score between demented group and the nondemented group due to the extremely small p value. In the result, we once confirmed that dementia can significantly affect the MMSE score of a person and therefore affect one's cognitive function.

4.1.2 Model Assumptions

The model assumptions need to be checked to ensure the accuracy of the model, Sphericity and Normality are checked. Sphericity is checked through mauchly's test. The result of Sphericity is 1.0 which means that the assumption of sphericity is met. However, the normality test result shows that for first visit and second visit, two p values are significantly small, 1.76e-13 and 5.43e-14, which means the normality assumption is violated.

4.1 Model 2: nWBV

Null hypothesis: 1, there is no significant difference in nWBV score across two visits for all patients. 2, there is no significant difference in nWBV score across all dementia groups. 3, The changes in nWBV score over two visits is the same for all groups.
Alternative hypothesis: 1, there is significant difference in nWBV score across two visits for all patients. 2, there is a significant difference in nWBV score across all dementia groups. 3, The changes in nWBV score over two visits differ across groups.

4.1.1 Model Result

The model result shows that the sum of squares for the interaction effect is extremely small. For group effect and Visit effect, we have detected smaller p values, 0.002 and <0.001, which mean that we reject the null hypothesis that there is no significant difference in nWBV score across two visits or across dimensional groups for all patients. However, the p value of the interaction effect, 0.171, is larger than the significant level which means that we fail to reject the null hypothesis that The changes in nWBV score over two visits is the same for all groups.

	SS	DF1	DF2	MS	F	P value
Group	0.034	2	140	0.017	6.000	0.002
Visit	0.007	1	140	0.007	97.523	<0.001
Interaction	<0.001	2	140	<0.001	1.786	0.171

The result of the post hoc test tells us that a significant difference in nWBV score can only be found between the nondemented group and demented group.

4.1.2 Model Assumptions

Similar to the previous model, we have to make sure the assumptions of sphericity and normality are met. The sphericity is met since the result of Mauchly's test is 1.0, same as the previous model. Thus, the sphericity assumption is met. The normality assumption is also met since both p values of first and second visit are high, 0.372 and 0.341. Thus the two assumptions are all met.

5. Power Analysis

I have made a power curve for a range of sample size from 3 to 100 and with effect size selected from 0.30, 0.50, 0.70,1.0 and 1.30. From the power curve we can see that as the sample size increase, the power tend to be closer to 1 for all effect size curve, but one thing different is how fast the power gets closer to 1 on different effect size, we can observe that as the effect size will be larger, the power is reaching to 1 more quickly.

5.1 Appropriate Sample size

Given that the power = 0.91, alpha = 0.5, and effect size = 0.7, the appropriate sample size is 46.

6. Conclusion

In Conclusion, we have found the MMSE scores are indeed affected by the existence of dementia and we have insight that the existence of dementia will result in more rapid decay in cognitive function. However, we can not conclude that the change over time of nWBV between patients with or without dementia is significantly different. Our result may be inaccurate due to violation of model assumptions.